

LOS ESTÁNDARES PARA LA EDICIÓN ELECTRONICA

Pedro Hípola. "Los estándares para la edición electrónica", en *Jornada sobre Publicación Electrónica, organizada por Socadi (Societat Catalana de Documentació i Informació), Barcelona, 28 de noviembre de 1996.*

http://www.ugr.es/~phipola/LOS_ESTANDARES_PARA_LA_EDICION_ELECTRONICA.pdf

Resumen:

En los últimos años se ha preparado una serie de normas destinadas a facilitar el intercambio de información electrónica de forma abierta. Se avanza así hacia la normalización del «documento electrónico». En este trabajo se analizan las diversas *tendencias* presentes en dichas normas.

Palabras clave:

Normalización, «documento electrónico», sistemas orientados al documento, SGML, ODA, Acrobat, Envoy, OpenDoc, OLE, DEN, ODMA, Shamrock, HTML.

LOS ESTÁNDARES PARA LA EDICIÓN ELECTRONICA

Pedro Hípola
Facultad de Biblioteconomía y Documentación
Universidad de Granada
18071 Granada
Tel.: 958-243938, fax: 243945
phipola@ugr.es

Información electrónica y normalización

El mundo de la edición electrónica necesita contar con *normas* que faciliten el trabajo de quienes están implicados en el proceso: autores, editores, distribuidores, fabricantes de hardware y de software, usuarios...

La mayor parte de las normas disponibles para la gestión de información electrónica sirven para resolver problemas informáticos en su más «bajo nivel»: compatibilidad de hardware, sistemas de codificación de caracteres, normas para las señales físicas en telecomunicaciones, etc. Sin embargo, es interesante observar la creciente importancia que van adquiriendo nuevas «familias» de normas aparecidas en los últimos años, destinadas a facilitar el intercambio de información electrónica de forma abierta, no sujeta a las especificaciones de un solo fabricante, pero todo ello a un nivel *superior*.

Algunas de estas normas, las menos pretenciosas, tratan de homogeneizar el formato y la estructura de los documentos que incluyen únicamente caracteres alfanuméricos. Las más completas abarcan todas las posibilidades de un entorno multimedia.

En los siguientes párrafos analizamos diversos sistemas encaminados a facilitar la normalización del «documento electrónico», intentando identificar las diversas *tendencias* presentes entre quienes están participando en el proceso normalizador.

La opción «semántica»

Una de las filosofías que pueden ser utilizadas a la hora de normalizar el «documento electrónico» es asignar a cada elemento del documento un *valor significativo* predeterminado.

Aunque no es una estructura específicamente diseñada para lo que normalmente se entiende por edición electrónica, podemos citar como ejemplo el estándar *ISO 2709*, la norma marco de las diversas versiones del formato Marc. Dentro de los registros Marc se puede reconocer el «valor semántico» del contenido de cada campo según cuál es el *designador de contenido* que lo identifica.

Salvando las distancias, algo similar podemos observar en el lenguaje *SGML* (*Standard Generalised Mark-up Language*), la norma más extendida dentro del mundo de la edición electrónica.

El lenguaje utiliza un conjunto de códigos que indican la estructura del documento. Sus diversos elementos (título principal, títulos secundarios, notas a pie de página, etc.) reciben una serie de codificaciones específicas. Esto hace que cada documento sea como un registro con sus respectivos campos.

Para crear un documento *SGML* es necesario hacer uso de una *Definición de Tipo de Documento (DTD)*, que especifica qué elementos pueden aparecer y qué reglas determinarán su presencia o ausencia.

Entre los más famosos *DTD* se encuentra el de la iniciativa *CALS* (*Computer-aided Acquisition and Logistic Support*), del *Departamento de Defensa* norteamericano, y la variante específica de *SGML* utilizada por los sistemas *WWW* (*World Wide Web*), denominada *HTML* (*HiperText Mark-up Language*).

Con cualquier procesador de textos es posible editar un fichero Ascii que incluya etiquetas *SGML*. Como esto puede resultar laborioso, existen editores (software) especiales para crear y manipular ficheros *SGML*. Pero incluso los procesadores de texto más populares (como es el caso de *WordPerfect* y de *Word*) han comenzado a dar soporte al formato.

Dentro de los sistemas que han optado por la caracterización semántica, merecen una mención especial los *sistemas EDI*.

Aunque no están dirigidos específicamente al mundo de la edición electrónica, vale la pena hacer una breve referencia a ellos por ser un ejemplo claro de lo que hemos denominado la «opción semántica», también por su éxito y porque pueden servir de modelo para desarrollar métodos de normalización documental (1).

Los sistemas EDI tratan de hacer posible el intercambio de pedidos, facturas, acuses de recibo, etc., entre los ordenadores de todo tipo de organizaciones. Con ello se facilita la automatización de determinadas tareas, eliminando esfuerzo humano, recursos económicos, tiempo, etc.

Las dos normas EDI más extendidas son *EDIFACT* en Europa y *ANSI X.12* en Estados Unidos. En el *American National Standards Institute (ANSI)* se ha estado trabajando para compatibilizar ambas normas.

Documentos compuestos «estáticos»

Otra tendencia la podemos observar en ciertos sistemas que permiten manipular documentos *compuestos «estáticos»*. Este tipo de documentos, también denominados «pasivos» (2), son los que integran sus distintos elementos perfectamente organizados de cara a poder obtener una presentación formal definitiva, en la pantalla o en el papel. Además pueden estar identificadas las relaciones lógicas entre sus diversas partes: gráficos, notas de pie de página, etc.

Entre estos sistemas podríamos incluir los *lenguajes de descripción de páginas (PDL)*, el más importante de los cuales es *PostScript*, de *Adobe*, y formatos «enriquecidos», como *TEX* (utilizado, por ejemplo, con el editor *LaTEX*) y *RTF* (*Microsoft*).

Otros sistemas, más sofisticados, están concebidos para facilitar el *intercambio* de ficheros compuestos entre plataformas informáticas variadas. En este

grupo hay que citar la norma ODA, *PDF-Acrobat* (Adobe), *Replica* (Farallon Computing), *Common Ground* (No Hands), *Envoy* (Novell-WordPerfect), *Bento* (Apple) y *CDA* (Digital).

ODA (*Open document architecture*) es un estándar (ISO 8613) poco utilizado, cuyo objetivo es conseguir que «los documentos, tanto textuales como gráficos, puedan ser transferidos, con todos sus atributos intactos, de un sistema a otro, para que luego puedan ser editados, procesados, almacenados, impresos y transmitidos».

Ofrece todo un marco global para la representación de documentos que incluyan materiales textuales y gráficos. Consta de una estructura física en bloques -título, párrafos, índice de materias, palabras clave, referencias externas, resumen, notas a pie de página...- y de una estructura lógica -páginas, cuadros, etc.-.

La norma ha recibido un mayor impulso desde 1991, fecha en la que se creó el *ODA Consortium*, constituido por *Bull*, *DEC*, *IBM*, *ICL*, *Siemens Nixdorf* y *Unisys*. Después se agregaron otros, como *WordPerfect*. Sin embargo, ODA no está teniendo mucho éxito.

Ante el posible fracaso de ODA, las empresas especializadas en software para la gestión de documentos electrónicos están enzarzadas en una carrera para ver quién consigue imponer como norma su propio formato de intercambio.

Adobe Systems Inc. presentó en 1993 *Acrobat*, un formato con el que quiere facilitar el intercambio de documentos electrónicos entre sistemas software y hardware que son incompatibles entre sí.

La misma empresa californiana fue quien lanzó en 1985 el lenguaje de descripción de páginas *PostScript*, que ha sido hasta ahora el principal estándar de facto para el intercambio de documentos electrónicos ya formateados. *PostScript* ha servido fundamentalmente para que diversos sistemas informáticos puedan imprimir los documentos creados por equipos de diferentes fabricantes.

El formato propio de *Acrobat* se denomina *PDF* (*Portable Document Format*), que puede incluir no sólo textos y gráficos, sino también anotaciones, enlaces hipertexto...

La versión 2 de *Acrobat*, recientemente presentada, incorpora importantes mejoras. La más notable, quizá, es que se ofrece la posibilidad de establecer enlaces con documentos y aplicaciones externos, utilizando el sistema de *URL* (*Universal Resource Locator*), propio de *WWW*. Esta versión se entrega también con un sistema de indexación y recuperación de los textos contenidos en los ficheros PDF.

Adobe está trabajando, con la ayuda de la empresa *Avalanche*, para que *Acrobat* pueda incluir codificación *SGML*. De esta manera el producto de *Adobe* asumiría las prestaciones que ofrece la norma ISO.

Existen varios productos que en cierta manera compiten con *Acrobat*: *Replica* (de *Farallon Computing*), *Common Ground* (*No Hands*), *Envoy* (*Novell-WordPerfect*) y *Bento* (*Apple*). Por la importancia de las casas que los han desarrollado, los dos últimos tendrán posiblemente bastante acogida. En el caso de *Envoy*, el producto es aún muy reciente y no está muy desarrollado, pero la posición predominante del procesador de textos *WordPerfect* a nivel mundial augura una amplia difusión de *Envoy*.

Documentos compuestos «dinámicos»

Una de las tendencias más destacables de la informática personal durante estos últimos años es el creciente interés por desarrollar lo que se ha denominado «sistemas orientados al documento», entornos operativos que ponen el énfasis en los documentos más que en las aplicaciones. El documento es el centro, y todo gira en torno a él. Éste puede integrar en su seno una serie de elementos -textos, gráficos, tablas, imágenes, sonido...- creados por diferentes aplicaciones y que permanecen ligados a ellas.

Con estos sistemas se pueden gestionar documentos *compuestos* «*dinámicos*», también denominados «*vivos*» (2). Cada parte del documento puede más adelante ser modificada utilizando una aplicación distinta, y conservando siempre su relación con el resto de las partes.

Desde el punto de vista orgánico, la información de los documentos se almacena en diferentes ficheros, lo que permite una mayor flexibilidad al sistema: cada aplicación puede operar sobre los correspondientes ficheros.

Entre los sistemas de gestión de documentos compuestos dinámicos destacan *OpenDoc* y *OLE*.

OpenDoc lo está promoviendo un consorcio de empresas liderado por *IBM*, *Apple* y *Novell-WordPerfect*. En el consorcio se encuentran, además de las tres compañías citadas, *Adobe*, *Hewlett Packard*, *Lotus*, *Oracle*, *Taligent* y *Xerox*.

El consorcio ha creado los *CI Labs (Component Integration Laboratories)*, donde se están desarrollando las especificaciones del sistema. La tecnología de almacenamiento de datos utilizada es la de *Bento*, formato de *Apple*. *IBM* aporta el *System Object Model* para controlar la compatibilidad binaria de los «objetos». Y *WordPerfect* contribuye con su experiencia en integración de productos dentro de *Windows*.

OpenDoc se fundamenta en las técnicas -tan de moda entre los programadores actuales- de «orientación a objetos». En concreto hace uso de la tecnología *Corba (Common Object Request Broker Architecture)*. Cada componente de un documento *OpenDoc* es un objeto reprocesable por la aplicación que lo creó.

En el listado de empresas comprometidas con *OpenDoc* no está *Microsoft*, pues este gigante informático del software defiende «en solitario» su propio sistema: *OLE (Object Linking and Embedding)*, que está concebido únicamente para ser explotado dentro de los entornos operativos que está desarrollando la misma empresa.

Todos los principales fabricantes de software están ajustando con prisa sus productos a las exigencias de la versión 2.0 de *OLE*. De hecho, entre los sistemas que se van a ajustar a la tecnología *OLE* está... ¡*OpenDoc*!

Arquitectura cliente/servidor

Dentro del esquema *cliente/servidor* encontramos otra de las tendencias que queremos ahora analizar. Existe una buena cantidad de productos, la mayoría poco conocidos en nuestro país, que son capaces de gestionar documentos cuyas partes están distribuidas por una red. Lo normal es que en el servidor se encuentre la aplicación central, que, además de albergar los ficheros principales, puede tener incorporado un sistema de índices y un motor de recuperación de información. Por su parte, cada aplicación cliente cuenta con su propio interface para acceder al servidor (3).

Algunos de estos productos son: *Asksam* (*Asksam*), *HotDocs* (*CapSoft*), *Documentum* (*Documentum*), *Excalibur* (*Excalibur*), *Visual Document Library* (*IBM*), *Relational Document Manager* (*Interleaf*), *Notes* (*Lotus*), *Document Direct* (*Mobius*), *ISYS* (*Idyssey*), *PC Docs Open* (*PC Docs*), *DocStor* (*Salix Systems*), *SoftSolutions* (*SoftSolutions*), *Topic* (*Verity*), *Open/Profound* (*Wang*), *Visual Recall* (*XSoft*).

En julio de 1994 *Novell* y *Xerox* hicieron público un acuerdo para poner en marcha los servicios *DEN* (*Document Enabled Networking*), cuyo objetivo es proporcionar un estándar abierto que permita a los fabricantes de software crear documentos compuestos que sean gestionables por diversas aplicaciones en un entorno de red. Se trata, en definitiva, de que cualquier aplicación cliente se pueda «entender» con cualquier aplicación servidor.

Por otra parte, otro grupo de empresas ha desarrollado *ODMA* (*Open Document Management*), un API (interface de programación de aplicaciones) que permite a las aplicaciones cliente entenderse con las aplicaciones de los servidores. Las prestaciones de *ODMA* son similares a las que se ha propuesto ofrecer *DEN*. Por otra parte, la presencia del gigante *Novell* en ambos proyectos ha hecho pensar que en el futuro *ODMA* puede pasar a formar parte de los servicios *DEN* (4). En cualquier caso, pocos auguran continuidad a *ODMA*.

Citemos, por último, la *Shamrock Document Management Coalition*, que, desde mediados de 1993, pretende conseguir la total interoperatividad entre todos

los sistemas. Para ello se está definiendo un conjunto de llamadas y servicios comunes que han de componer la arquitectura *Enterprise Document Management (EDM)*.

El paradigma hipertexto/hipermedia

Como última tendencia nos referiremos a los sistemas que permiten ofrecer prestaciones *hipertexto/hipermedia*. Algunos de los sistemas ya analizados en párrafos anteriores lo hacen. Es el caso de *SGML*, *PDF-Acrobat*, *Envoy*... Pero quizá el ejemplo más llamativo es el de *HTML (Hypertext Markup Language)*, el lenguaje de etiquetas que hace posible que exista *WWW (World Wide Web)*, la «telaraña mundial».

Este sistema aglutina todas las tendencias que hemos venido analizando: documentos compuestos dinámicos, capacidad multimedia, arquitectura cliente/servidor, hipertexto e hipermedia, todo ello con gran sencillez. Un usuario medio es capaz de elaborar documentos completos *HTML* tras sólo un par de horas de aprendizaje.

Con un cliente *WWW* podemos conectarnos a un servidor *WWW*. Éste nos presenta un documento inicial en pantalla, la «home page». Lo normal es que esta página incluya diversos elementos resaltados de alguna manera, que son los que contienen enlaces con otras partes del mismo documento, con otros documentos, gráficos, etc.

Estos nuevos materiales pueden residir físicamente en el mismo servidor o en otro servidor situado en algún punto de la red (en diversas partes del planeta, si estamos en Internet). El software gestiona la conexión necesaria de forma automática y transparente. Y sin que apenas nos demos cuenta nos desplazamos por los documentos existentes en los lugares más insospechados.

REFERENCIAS

- (1) **Pedro Hípola y Félix de Moya**, "Proyectos EDI y normalización documental", en *Revista española de Documentación científica*, octubre-diciembre 1991, 4, p. 408-419.
- (2) **Isidre Canals**, "Los documentos compuestos vivos, base de los futuros sistemas orientados al documento", en *Information World en Español*, junio 1994, 25, p. 14-16.
- (3) **Linda Musthaler**, "A tall order for document managers", en *Network World*, 18 de julio de 1994, p. 35-44.
- (4) **Andy Reinhart**, "Managing the new document", en *Byte*, agosto 1994, p. 91-104.

Información electrónica
y normalización
Problemas básicos

- Crecimiento industria información electrónica
- Productos «aislados» / normalizados
- Normalización a «bajo nivel»:
 - * compatibilidad de hardware
 - * codificación de caracteres
 - * señales físicas en telecomunicaciones
- Necesidad de niveles superiores
- Materiales mono-media / multimedia
 - * textos: ASCII
 - * imágenes fijas: FAX (CCITT), TIFF, PCX, JPEG
 - * imágenes en movimiento: MPEG
 - * sonido: MIDI, WAVE

Normalización del
«documento electrónico»
Tendencias

La opción «semántica»:

Documentos compuestos «estáticos»

Los sistemas «orientados al documento»

Documentos compuestos «dinámicos»
(Document Management Systems)

El paradigma hipertexto/hipermedia

Arquitectura cliente/servidor

Normalización del
«documento electrónico»
Tendencias

- La opción «semántica»:

- * ISO 2709 (...)
- * SGML (Standard Generalised Mark-up Lang.)
- * HTML (...)
- * EDI (Electronic Data Interchange)
 - > Edifact
 - > Ansi X.12
- * ODA/ODIF (Office document architecture/Office document interchange format)

- Documentos compuestos «estáticos»

Monomedia

- * PDL: PostScript (Adobe)
- * TEX, RTF - Rich Text Format (Microsoft), PDF-Acrobat (Adobe), Replica (Farallon Comp.), No Hands (Common Ground), Envoy (Word Perfect), Bento (Apple)

Multimedia

- * ODA/ODIF
- * CDA (DEC)

Normalización del
«documento electrónico»
Tendencias

- Los sistemas «orientados al documento»
- Documentos compuestos «dinámicos»
(Document Management Systems)
 - * Object Linking and Embedding (Microsoft)
 - * OpenDoc
- El paradigma hipertexto/hipermedia
 - * SGML
 - * WWW-HTML (HyperText Mark-up Language)
 - * PDF-Acrobat, Envoy...

Normalización del
«documento electrónico»
Tendencias

- Arquitectura cliente/servidor

- * WWW-HTML

- * Asksam (Asksam), HotDocs (CapSoft), Documentum (Documentum), Excalibur (Excalibur), Visual Document Library (IBM), Relational Document Manager (Interleaf), Notes (Lotus), Document Direct (Mobius), ISYS (Idyssey), PC Docs Open (PC DOCS), DocStor (Salix Systems), SoftSolutions (SoftSolutions), Topic (Verity), Open/Profound (Wang), Visual Recall (XSoft)

- DEN (Document Enabled Networking) API (Novell + Xerox)

- ODMA (Open Document Management API): Adobe, Andersen, Autodesk, Borland, Documentum, Interleaf,

Microsoft, Novell, Oracle, PC DOCS, SoftSolutions, Sybase, Wordperfect, Xsoft

- Shamrock Document Management Coalition: Adobe, Aetna, Andersen, Coca-Cola, Documentum, EDS, Frame Technology, Hewlett-Packard, IBM, Interleaf, Merck, Microsoft, PC DOCS, Sybase, Verity, ViewStar, Wang, Xsoft

-> total interoperatividad

SGML

(Standard Generalised Mark-up Language)

- ISO 8879

- códigos (mark-up): título principal
 títulos secundarios
 notas pie de página
 (...)

- documento -> registro (campos)

EDI

«transferencia directa entre ordenadores, a través de medios electrónicos, de datos de negocios estructurados, esto es, la transferencia de 'documentación' de negocios sin papeles»

(International Chamber of Commerce, *Uniform rules of conduct for interchange of trade data by teletransmission*, 1988)

EDIFACT: 1) vocabulario Tded (trade data elements directory), ISO 7372 y EN 29732

2) reglas de sintaxis al nivel de aplicación de OSI (open systems interconnection), ISO 9735 y EN 29735

3) un directorio de segmentos normalizados: nombre, dirección, tipo de intercambio...

4) mensajes normalizados: orden de pedido, factura, aviso de envío, declaración de aduanas...

5) códigos: países, monedas, modos de transporte, condiciones de pago...

ANSI X.12:

ACROBAT

Adobe Systems

-> intercambio de documentos electrónicos entre sistemas software y hardware «incompatibles»

- 1988: PostScript

- 1993: Acrobat

- PDF (Portable Document Format):

- * texto + gráficos
- * anotaciones
- * textos buscables
- * enlaces hipertexto

- Gran capacidad de compresión

<- fax grupo 4 (CCITT)

<- JPEG

- Éxito (?):

Cajun (CD-ROM Acrobat Journals Using Networks)

<- John Wiley & Sons y Chapman & Hall

Dataware

Presupuesto USA

(...)

- sistema «propietario»

ACROBAT

Adobe Systems

- * Acrobat Exchange:
 - generación ficheros PDF
 - presentación en pantalla e impresora
 - manipulación (cortar y pegar, anotaciones, búsquedas, navegación...
- * Acrobat Reader:
 - presentación en pantalla e impresora
- * Acrobat Distiller:
 - traducción PostScript -> PDF
 - funciones de zoom
- * Acrobat Search:
 - búsquedas

* Acrobat Catalog:
índices

REPLICA

(Farallon Computing)

- <- tecnología TrueType (caracteres escalables)
- Funciones recortar y pegar
- No anotaciones
- Módulo «embebido» en el documento
- sistema «propietario»

NO HANDS

(Common Ground)

- Módulo «embebido»
- Recortar y pegar
- Búsquedas de cadenas de caracteres
- sistema «propietario»

ENVOY

(Word Perfect)

* Envoy Driver: creación de ficheros

* Envoy Viewer: presentación, manipulación e impresión

Distribución gratuita

- Hipertexto

- No búsquedas

- sistema «propietario»

OPENDOC

Adobe, Apple, Hewlett Packard, IBM, Lotus, Novell-Word Perfect, Oracle. Taligent, Xerox

-> CI Labs (Component Integration Labs)

- Bento (Apple) -> almacenamiento de datos
- System Object Model (IBM) -> compatibilidad de objetos
- Word Perfect -> integración en Windows
- Información multimedia
- Edición «in place»
 - > independencia de la aplicación creadora
- Integración multifichero
- Soporte multiplataforma
- Tecnología Corba (Common Object Request Based Architecture)

- Documento <- partes <- partes (...)

* Content containers (estructuras de datos)

* Part editors (programas independientes)

* Frames (delimitadores de ámbito)

* Part handlers (aplicaciones): editores y visualizadores

WWW
(World Wide Web)

ODA/ODIF: ISO 8613